CLC **Genomics** Server

User manual

Manual for

CLC Genomics Server 4.5

Windows, Mac OS X and Linux

June 12, 2012

This software is for research purposes only.

CLC bio
Finlandsgade 10-12
DK-8200 Aarhus N
Denmark

Contents

1 Introduction **4**

2 Using the server from a CLC Workbench **6**

2.1 Installing relevant plug-ins in the Workbench 6

2.2 Log in to the server 6

2.3 Browsing and searching data from the workbench 7

2.4 Running analyses on the server 8

2.5 Accessing files on, and writing to, areas of the server filesystem 9

2.6 Monitoring processes 11

3 Using a web browser as client **13**

3.1 Browsing and searching data 13

3.2 Upload and download 14

Chapter 1

Introduction

The *CLC Server* is the central part of CLC bio's enterprise solutions. You can see an overview of the server solution in figure 1.1).

The basic idea is that you can have the Server store data and run jobs centrally and thereby off load personal and typically smaller computers. For the user, the difference between working just with a Workbench and working with a Workbench *and* a server is very small. All the mechanisms for managing data, using the tools and visualizing the data are the same.

This user manual is intended to show how to take advantage of the server as a user. For information about administrating the server, please see the Administrator Manual.

The *CLC Genomics Server* is shipped with the following tools and analyses that can be started from the *CLC Genomics Workbench*:

- Import
- Trimming
- Sequencing Data QC Report
- Map reads to reference
- De novo assembly
- All tools for re-sequencing analysis
- Download genomes
- All tools for track management and filtering
- Create detailed mapping report
- ChIP-Seq
- RNA-Seq
- Extract and count small RNA
- Annotate small RNA

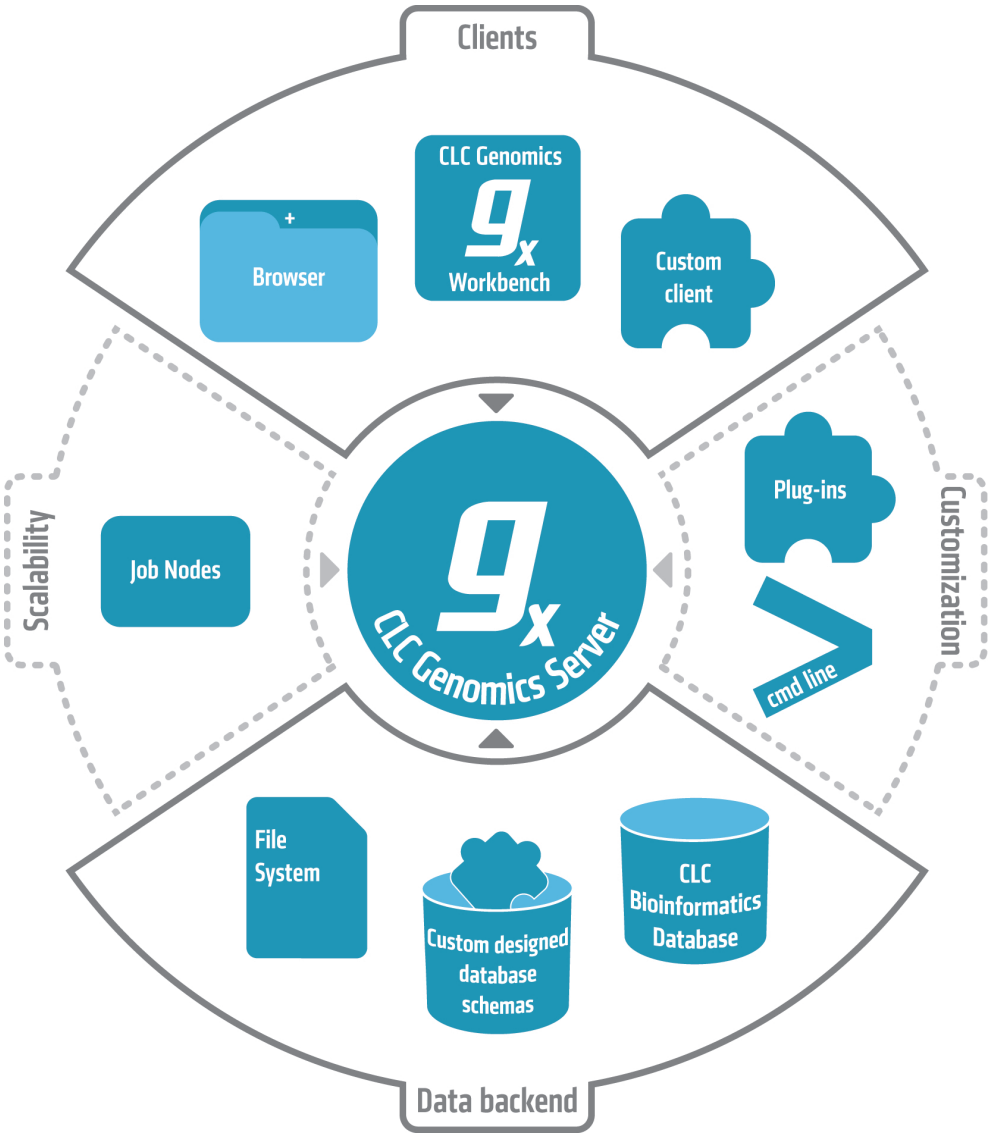

Figure 1.1: *An overview of the server solution from CLC bio. Note that not all features are included with all license models.*

- Process tagged sequences (de-multiplexing)
- External applications (integrating with 3rd party programs on the server)
- BLAST (creation of databases, download of databases and running BLAST against local data or NCBI)
- Find primer binding sites
- Annotate secondary peaks (Sanger data)
- Extract sequences
- Additional alignments (plug-in)

The functionality of the *CLC Genomics Server* can be extended by installation of Server plug-ins. The available plug-ins can be found at http://www.clcbio.com/server_plugins.

Chapter 2

Using the server from a CLC Workbench

2.1 Installing relevant plug-ins in the Workbench

In order to use the *CLC Genomics Server* from a CLC Workbench, you need to install a plug-in to the Workbench. If you connect to the server from a *CLC Genomics Workbench*, you should install the CLC Genomics Server Plug-in. If you are using e.g. *CLC Main Workbench*, you should install the generic CLC Server Plug-in. Both plug-ins allow you to log in to the server and access data from the server data locations, but the CLC Genomics Server Plug-in also includes the tools to access the analyses on the server (e.g. assembly, RNA-Seq).

The plug-ins can be found on the same web page as the server installer or at http://www.clcbio.com/plugins.

Plug-ins are installed using the plug-in manager[1]:

Help in the Menu Bar | **Install Plug-ins (** **)**

or **Plug-ins (** **) in the Toolbar**

Install the plug-in by clicking the **Install from File** button at the bottom of the dialog. This will open a dialog where you can browse for the plug-in. The plug-in file is provided by CLC bio on the page where you downloaded the server installation.

You need to restart the Workbench before the plug-in is ready for use.

2.2 Log in to the server

Once the plug-in is installed, log in to the server:

File | **CLC Server Login (** **)**

That will bring up a login dialog as shown in figure 2.1).

The first time you log in to the server, you have to expand the dialog by clicking **Advanced**. That will allow you to enter the host and port for the server as shown in figure 2.2).

In addition you can choose to save user name and password and automatically log in to the

[1] In order to install plug-ins on Windows Vista, the Workbench must be run in administrator mode: Right-click the program shortcut and choose "Run as Administrator". Then follow the procedure described below.

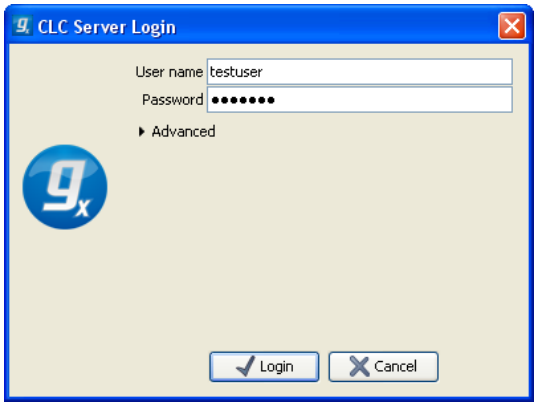

Figure 2.1: *Expand the login dialog by clicking Advanced.*

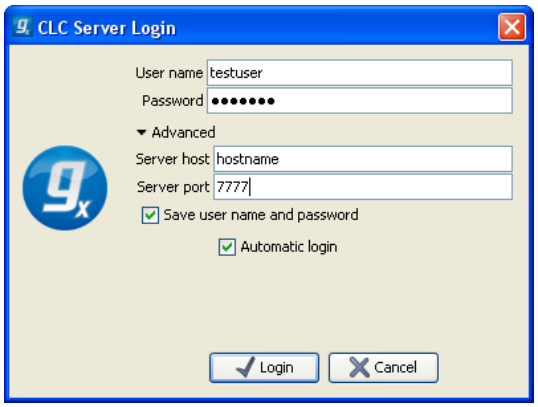

Figure 2.2: *Specifying host and port.*

server when the Workbench starts.

Note that you need to get information used to log in from your server administrator.

When you press **Login**, the Workbench connects to the server. You will see a progress bar in the login dialog. If the login is successful, the dialog will disappear, and you will be able to use the server as described below.

2.3 Browsing and searching data from the workbench

Once you have logged in, the data locations on the server are shown in the **Navigation Area** (see figure 2.3).

Once logged in, the server data locations can be used in the same way as local data locations. We refer to the users manual of the Workbench for information about using the **Navigation Area** (click the location and press **F1** on the keyboard to get context help).

You may also wish to have a look at the search chapter, *Searching your data*, in the workbench user manual (press F1 and look for *Searching your data*).

It is possible to have both local and server locations added at the same time. This means that you can work on e.g. temporary sequences located on your own computer and then when you have more complete results, you can drag the elements to a folder in the server location.

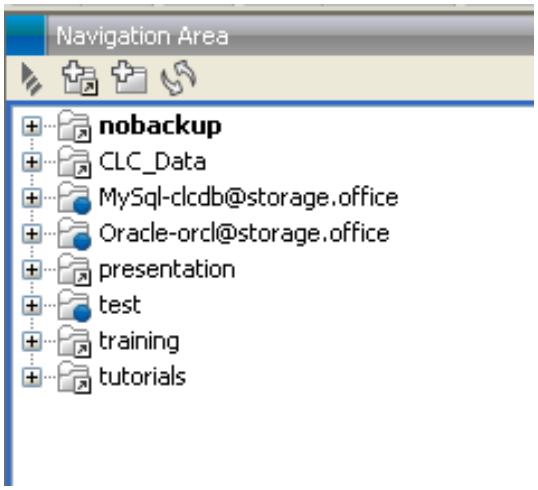

Figure 2.3: *Three server locations on the server appears in the Navigation Area. There are also five local file locations that exist in the Workbench independent of the server.*

2.4 Running analyses on the server

The tools available on the server are listed in the introduction (see section 1).

For more information on the tools, please see the manual for *CLC Genomics Workbench* at `http://www.clcbio.com/usermanuals` or click the **Help** (?) button in the dialogs that are shown when you run the tools.

The high-throughput sequencing data import tools are also available on the server as shown in figure 2.4.

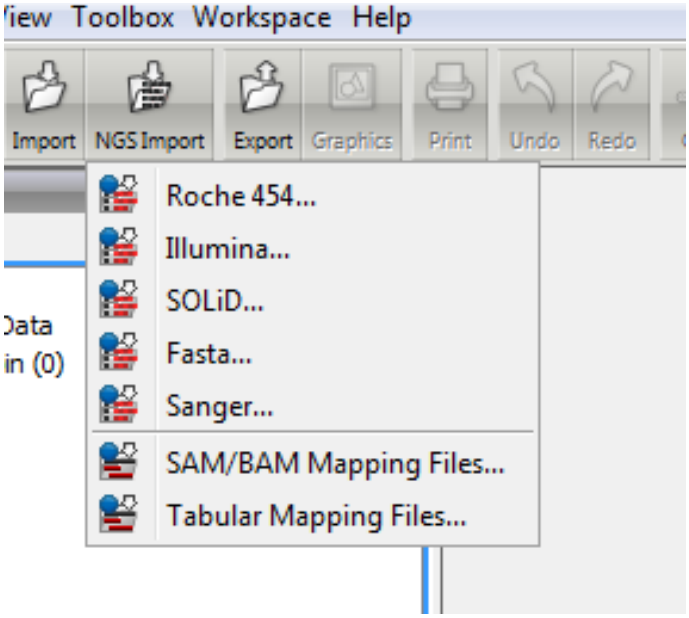

Figure 2.4: *Once logged in, the import tools on the server become active.*

When you run the analyses you will be faced with an initial dialog with two choices:

- **Workbench**. This will run the analysis on your computer.
- **Server**. This will run the analysis on the server.

An example of such a dialog is shown in figure 2.5.

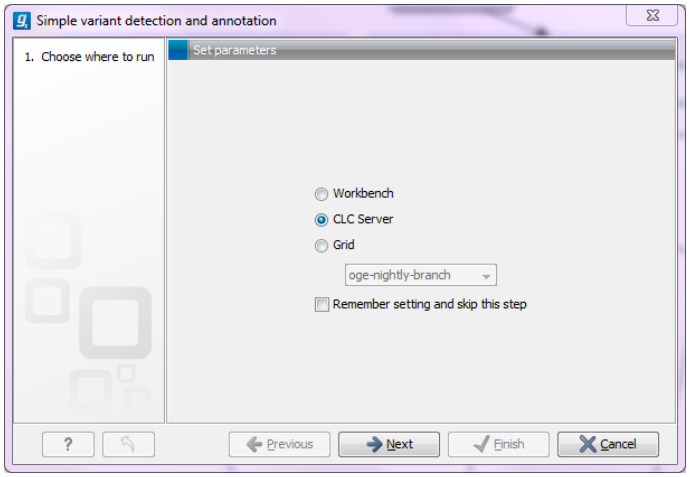

Figure 2.5: *Selecting where to run the analysis.*

Note that you can check the **Remember setting and skip this step** option to always use the selected option when the analysis dialog is shown. If you change your mind later on and want to switch, simply click **Previous** in the dialog and you will be able to change the setting.

Running the analyses on the server is identical to running the same analyses in the Workbench with three exceptions:

- You can only select data from the data locations connected to the server. This means that the Navigation Area is filtered to only show these locations.
- You have to save the result. In the last step, you can normally choose between **Open** and **Save**, but for the analyses performed on the server, you have to choose **Save**.
- When you click **Finish**, the analysis is performed on the server. This means that you can close the Workbench or disconnect the computer from the server, and the analysis will continue to run. When you log in after the analysis is done, you will get a notification in the Workbench. A special case is the tools for importing high-throughput sequencing data on the server (only for Genomics Workbench and Server). Depending on your choice, the first part of the import is uploading the data to the server and during this part, the Workbench has to maintain connection to the server. If you try close the Workbench during upload, you will get a warning dialog. The upload runs as a local process in **Processes** tab in the lower left corner, and when the upload is done, a new process is started that has a server icon and runs on the server.

If you need help on the analyses themselves, please click the **Help** (?) button in the dialog. That will take you to the relevant part of the user manual of the Workbench.

2.5 Accessing files on, and writing to, areas of the server filesystem

There are circumstance when it is beneficial to be able to interact with (non-CLC) files directly on your server filesystem. A common circumstance would be importing high-throughput sequencing data from folders where it is stored on the same system that your *CLC Genomics Server* is

running on. This could eliminate the need for each user to copy large sequence data files onto the machine their CLC Workbench is running on before importing the data into a *CLC Genomics Server* CLC server data area. Another example is if you wished to export data from CLC format to other formats and save those files on your server machine's filesystem (as opposed to saving the files onto the system your Workbench is running on).

From the administrator's point of view, this is about configuring folders that are safe for the *CLC Genomics Server* to read and write to on the server machine' system.

This means that users logged into the *CLC Genomics Server* from their Workbench will be able to access files in that area, and potentially write files to that area. Note that the *CLC Genomics Server* will be accessing the file system as the *user running the server process* - not as the user logged into the Workbench. This means that you should be careful when opening access to the server filesystem in this way. Certainly, only folders that do not contain sensitive information should be added.

Folders to be added for this type of access are configured in the web administration interface in the **Main configuration** tab under **Import/export directories** (see figure 2.6). (Earlier versions of the server software referred to these locations as High-throughput sequencing data import locations.)

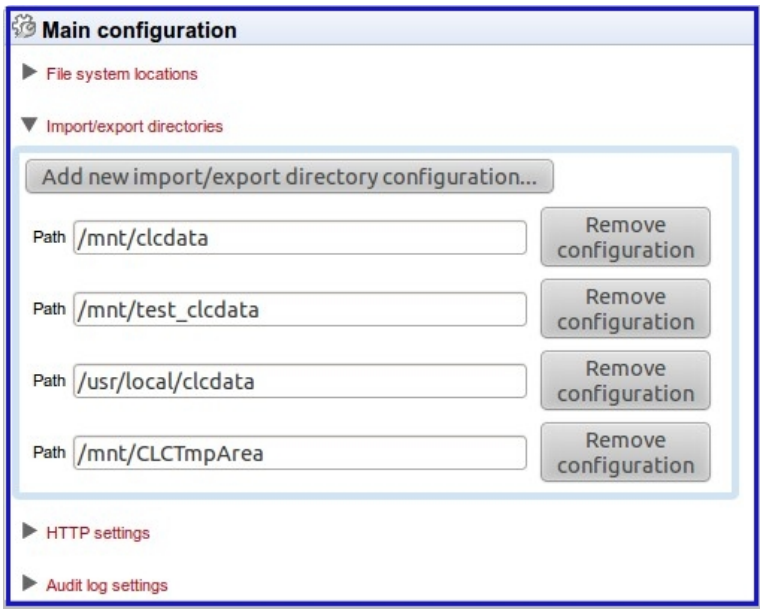

Figure 2.6: *Defining source folders that should be available for browsing from the Workbench.*

Press the **Add new import/export directory location** button to specify a path to a folder on the server. This folder and all its subfolders will then be available for browsing in the Workbench for certain activities (e.g. importing data functions).

For example, in the case where a directory has been set up an Import/export location, and a user, logged into the *CLC Genomics Server* via their CLC Workbench wishes to import some high throughput sequencing data, they would see an option like that shown in figure 2.7.

On my local disk or a place I have access to means that the user will be able to select files from the file system of the machine their CLC Workbench is installed on. These files will then be transferred over the network to the server and placed as temporary files for importing. If the user chooses instead the option *On the server or a place the server has access to*, the user is

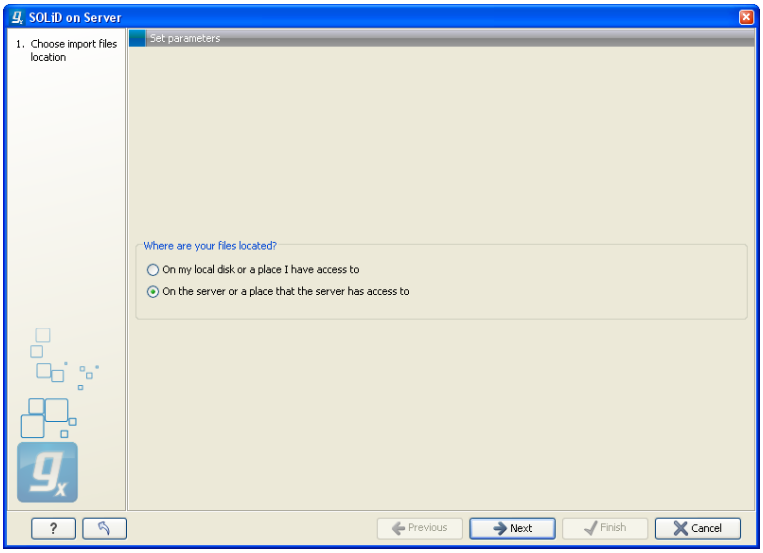
SOLiD on Server
1. Choose import files location
Set parameters
Where are your files located?
On my local disk or a place I have access to
On the server or a place that the server has access to
Previous
Next
Finish
Cancel

Figure 2.7: *Deciding source for high-throughput sequencing data files.*

presented with a file browser for the selected parts of the server file system that the administator has configured as an Import/export location. See figure 2.8.

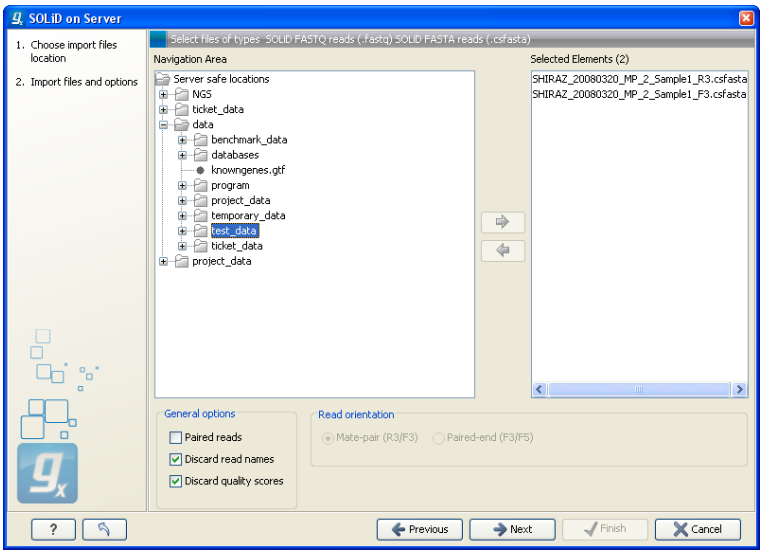

Figure 2.8: *Selecting files on server file system.*

Note: Import/Export locations should NOT be set to subfolders of any defined CLC file or data locations. CLC file and data locations should be used for CLC data, and data should only be added or removed from these areas by CLC tools. By definition, an Import/Export folder is meant for holding non-CLC data, for example, sequencing data that will be imported, data that you export from the Genomics Server, or blast databases. Note that your server administrator needs to configure the server to import files directly from the server file system.

2.6 Monitoring processes

Just as for processes running locally, you will be able to monitor process running on the server. Click the **Processes** tab next to the **Toolbox** and you will see the list of processes (see figure 2.9).

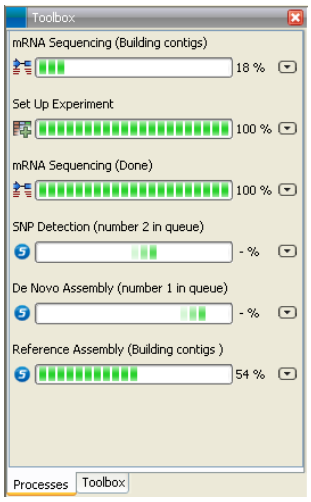

Figure 2.9: *Monitoring processes.*

Note that the processes running on the server have a server icon ( ) or ( ) whereas the processes running locally have specific icons like ( ) and ( ). You can also see that two of the server processes are queued.

Contrary to processes running locally, running server processes will appear even if you restart the Workbench (and log in to the server). Server processes that are finished when you close the Workbench will not be shown when you start the Workbench again.

Chapter 3

Using a web browser as client

Besides using the CLC Workbench as client, you also have access to the server from the web interface. The web interface gives you access to browsing and searching data and up- and downloading data.

Simply type the address of the server into your browser, follow by the port number and you will see the login dialog (see figure 3.1) (you will need to get information from your server administrator about the server address).

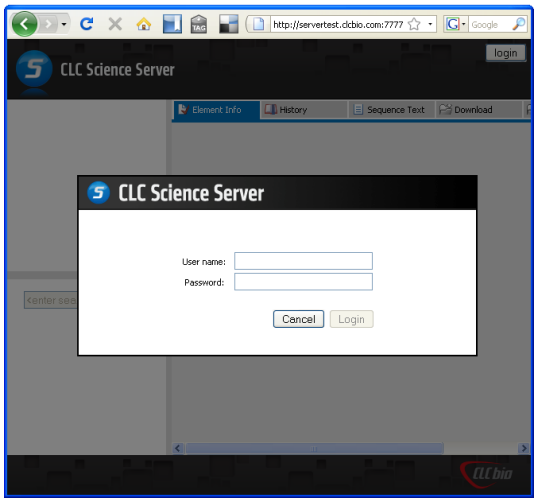

Figure 3.1: *The web interface of the server.*

3.1 Browsing and searching data

To the left, you will see the data locations connected to the server. You can browse the folder hierarchy of the data location. When you click an element in the tree, a number of options are available:

- Click the **Element Info** ( ) tab to see the properties of this element. Note that you can edit the information in this view.
- Click the **History** ( ) tab to see the history of this element.

- Click the **Sequence Text** ( ) tab to see a textual representation of this element (only works for sequences)

An example of a protein sequence in the text view is shown in figure 3.2.

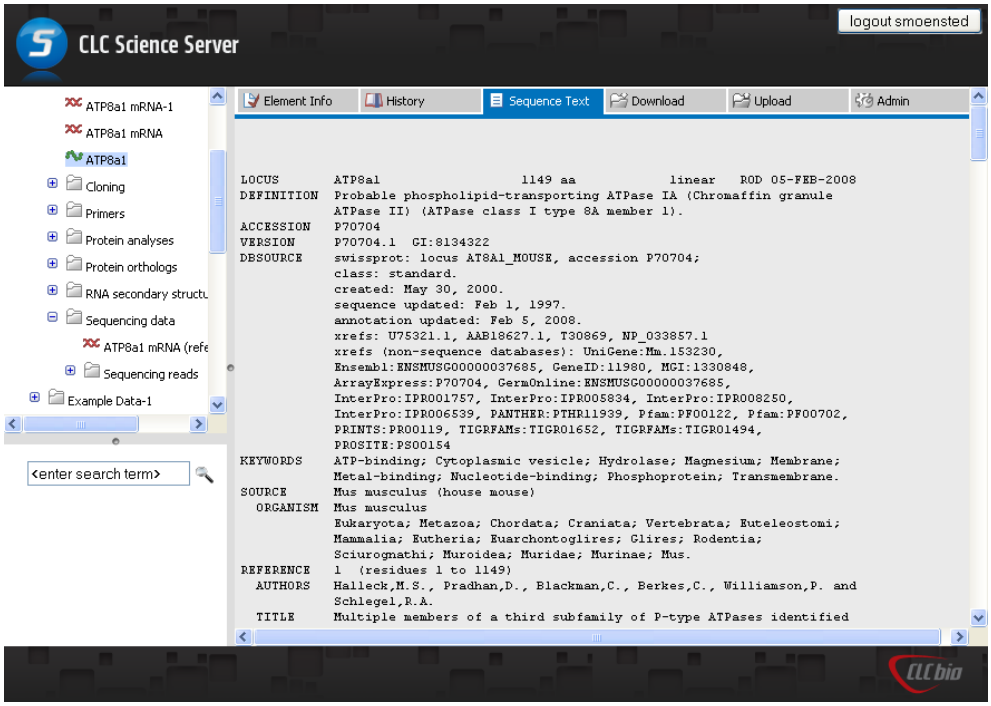
CLC Science Server
logout smoensted
ATP8a1 mRNA-1
ATP8a1 mRNA
ATP8a1
Cloning
Primers
Protein analyses
Protein orthologs
RNA secondary structu
Sequencing data
ATP8a1 mRNA (refe
Sequencing reads
Example Data-1
<enter search term>
Element Info
History
Sequence Text
Download
Upload
Admin
LOCUS ATP8a1 1149 aa linear ROD 05-FEB-2008
DEFINITION Probable phospholipid-transporting ATPase IA (Chromaffin granule ATPase II) (ATPase class I type 8A member 1).
ACCESSION P70704
VERSION P70704.1 GI:8134322
DBSOURCE swissprot: locus AT8A1_MOUSE, accession P70704;
class: standard.
created: May 30, 2000.
sequence updated: Feb 1, 1997.
annotation updated: Feb 5, 2008.
xrefs: U75321.1, AAB18627.1, T30869, NP_033857.1
xrefs (non-sequence databases): UniGene:Mm.153230, Ensembl:ENSMUSG00000037685, GeneID:11980, MGI:1330848, ArrayExpress:P70704, GermOnline:ENSMUSG00000037685, InterPro:IPR001757, InterPro:IPR005834, InterPro:IPR008250, InterPro:IPR006539, PANTHER:PTHR11939, Pfam:PF00122, Pfam:PF00702, PRINTS:PR00119, TIGRFAMs:TIGR01652, TIGRFAMs:TIGR01494, PROSITE:PS00154
KEYWORDS ATP-binding; Cytoplasmic vesicle; Hydrolase; Magnesium; Membrane; Metal-binding; Nucleotide-binding; Phosphoprotein; Transmembrane.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muroidea; Muridae; Murinae; Mus.
REFERENCE 1 (residues 1 to 1149)
AUTHORS Halleck,M.S., Pradhan,D., Blackman,C., Berkes,C., Williamson,P. and Schlegel,R.A.
TITLE Multiple members of a third subfamily of P-type ATPases identified
CLC bio

Figure 3.2: *Inspecting the text view of a protein sequence.*

Note that these views are a subset of the views that you find in the CLC Workbench.

3.2 Upload and download

If you wish to download one of the elements from the server, simply select it and click **Download** ( ). That will present a list of file formats as shown in figure 3.3.

Clicking one of the buttons will start the download of the sequence in the selected format.

You can also put data into the server data location: select the folder where you want the data to be, and click **Upload** ( ) (see figure 3.4).

Click the **Browse** button to select a file and click **Upload File**. The server will automatically recognize the file format and interpret the file. If this does not work as expected, you can try to select the file format in the drop-down list above the Upload File button (where it says "Automatic" in figure 3.4).

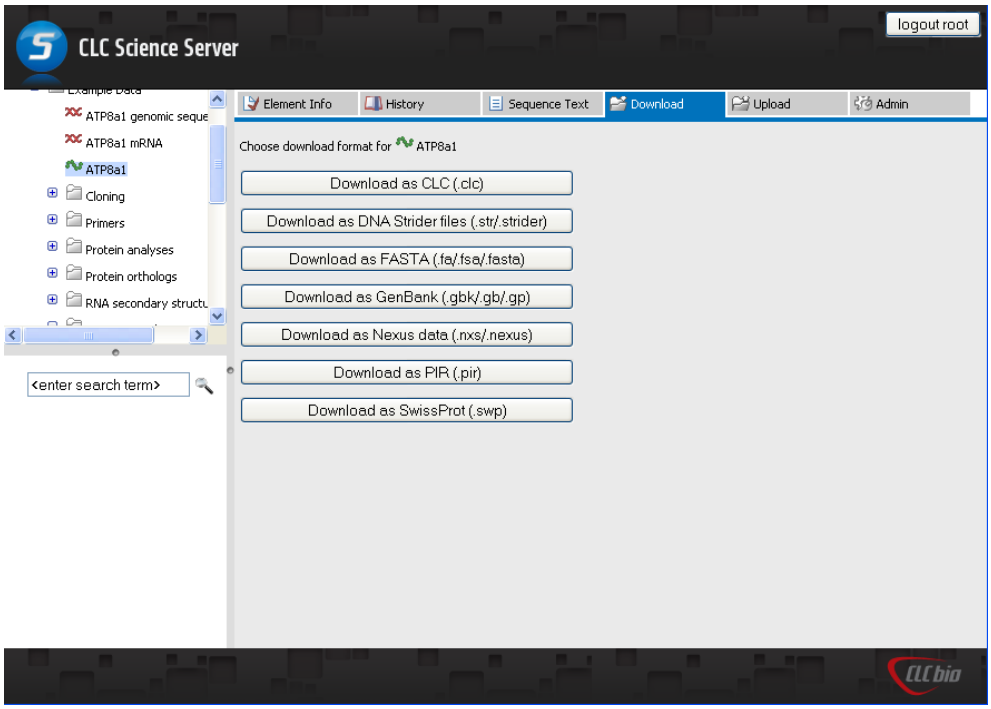

Figure 3.3: *Downloading a sequence from the server.*

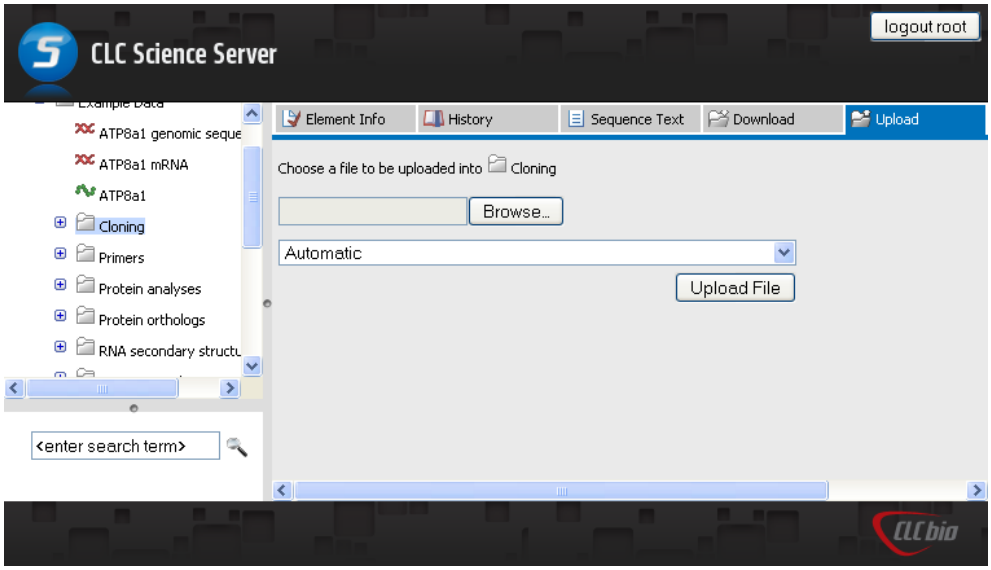

Figure 3.4: *Uploading sequences to the server.*